Generalized Do-Calculus Without Graphs

Benjamin Heymann, Michel de Lara, Jean-Philippe Chancelier

Generalized Do-Calculus Without Graphs

Benjamin Heymann[*], Michel De Lara[†], Jean-Philippe Chancelier[†]

July 30, 2020

Abstract

Inferring the potential consequences of an unobserved event is a fundamental scientific question. To this end, Pearl's celebrated do-calculus provides a set of inference rules to derive an interventional probability from an observational probability. In this framework, the primitive causal relations are encoded on a Directed Acyclic Graph (DAG), which can be limitative for some applications. In this paper, we capture causality without reference to a graph and we extend the rules of do-calculus to systems that do not possess a fixed causal ordering. For this purpose, we introduce a new framework which relies on the theory developed by Witsenhausen for multi-agent stochastic control. The mapping from graphs to so called Witsenhausen's intrinsic model is natural: the primitives of the problem are the agents' information fields; the random variables are synthesized by the agents whose strategies encode the informational constraints. All in all, our framework offers a richer language than DAGs and provides a generalized do-calculus.

1 Introduction

1.1 Causality and do-calculus

As the world shifts toward more and more data-driven decision making, causal inference is taking more space in applied sciences, statistics and machine learning. This is because it allows for better, more robust decision making, and provides a way to interpret the data that goes beyond correlation [12]. For instance, causal inference provides a language and tools to describe and solve Simpson's paradox, which embodies the "correlation is not causation" principle as can be found in any "Statistics 101" basic course. The main concern in causal inference is to compute post-intervention probabilities distribution from pre-intervention data. For this purpose, graphical models are practical because they allow to represent assumptions easily and benefit from an extensive scientific literature.

In his seminal work, Pearl builds on graphical models [3] to introduce the so-called do-calculus. Several extensions to this do-calculus have been proposed recently [19, 9, 17, 2]. As asserted by Pearl, language is an important element in this research program [10]:

> Another ramification of the sharp distinction between associational and causal concepts is that any mathematical approach to causal analysis must acquire *new notation* for expressing causal relations – probability calculus is insufficient.

We completely concur with this idea. Moreover, our contribution is based on the work of another scholar, Witsenhausen, who developed a mathematical framework to capture the notion of causality and who wrote

> The main difficulty is one of notations.

[*]Criteo, Paris, France

[†]CERMICS, Ecole des Ponts, Marne-la-Vallée, France

Witsenhausen introduced the intrinsic model that turns the focus from dependencies represented by functions to dependencies represented by σ-fields and measurability constraints (this will be clarified thereafter) in multi-agent stochastic control problems.

Causal graphical models move the focus from joint probability distributions to functional dependencies thanks to the Structural Causal Model (SCM). It is our hope that the intrinsic model will move the focus from functional dependencies to informational relations, hence bringing a new, complementary view to the causal reasoning toolbox.

So, we introduce a general, unifying framework for causal inference that may be used for both recursive and non-recursive systems [5]. The cost for this generalization is a bit of abstraction: in what we propose, the structure is implicit, and there are no arrows. In particular, while DAG modeling does not rely directly on the notion of random variable but on the joint, pushforward probability distribution ([11], footnote 3 or [13] Appendix A), our approach requires to go back to the primitives of the probabilistic model: sample set, *sigma*-fields, measurable maps. This is why we depart from the usual presentation of causal inference papers.

1.2 Our contributions

We generalise causal modeling thanks to the notion of information fields. In this new abstraction, we redefine the notion of d-separation and provide a rigorous analysis of the model structure. Our main results are Theorem 9, which is a characterization of this extended d-separability notion, Theorem 10 which expresses that independence induces a factorization property of the solution map, and Theorem 12, which is a generalization of do-calculus, and which subsumes several recent results. All proofs and examples are provided in the companion paper.

2 The Witsenhausen's intrinsic model and formalism for causality

We present the so-called *Witsenhausen's intrinsic model*, followed by its formalism for causality.

2.1 The Witsenhausen's intrinsic model

As Witsenhausen's intrinsic model was introduced some five decades ago in the control community [20, 21], it may not be familiar to all readers. We provide tentative correspondences between Pearl's DAG and Witsenhausen's intrinsic model in Table 1.

	Pearl	*Witsenhausen*
Structure	DAG	Nature and agents decision sets, with their respective fields
Parent relation	$\rightarrow$	precedence relation
	node	agent
	edge	agents related by the precedence relation
Dependence	SCM	agents information fields
	functional relation	policy profiles measurable w.r.t. information fields
Resolution	induction	solution map
	random variable	policy profile composed with solution map
Intervention	Do operator	change of information fields
Causal ordering	fixed	existence depends on agents' information fields

Table 1: Correspondences between Pearl's DAG and Witsenhausen's intrinsic model

Witsenhausen used the language of σ-fields to handle the concept of information in control theory. We present it below using the more general notion of field. A *field* (resp. *σ-field*) over the set $\mathbb{D}$ is a subset $\mathcal{D} \subset 2^{\mathbb{D}}$, containing $\mathbb{D}$, and which is stable under complementation and union (resp. countable union). The trivial field over the set $\mathbb{D}$ is $\{\emptyset, \mathbb{D}\}$. The complete field over the set $\mathbb{D}$ is $2^{\mathbb{D}}$. When $\mathcal{D}' \subset \mathcal{D}$ are two fields over the set $\mathbb{D}$, we say that $\mathcal{D}'$ is a *subfield* of $\mathcal{D}$.

Definition 1. *(Adapted from [20, 21]) A* W-model *is a collection* $\big(\mathbb{A}, (\Omega, \mathcal{F}), (\mathbb{U}_a, \mathcal{U}_a)_{a\in\mathbb{A}}, (\mathcal{I}_a)_{a\in\mathbb{A}}\big)$*, where*

- $\mathbb{A}$ *is a finite set, whose elements are called* agents*;*
- Ω *is a set made of* states of Nature*;* $\mathcal{F}$ *is a field over* Ω*;*
- *for any* $a \in \mathbb{A}$*,* $\mathbb{U}_a$ *is a set, the* set of decisions *for agent* a*;* $\mathcal{U}_a$ *is a field over* $\mathbb{U}_a$*;*
- *for any* $a \in \mathbb{A}$*,* $\mathcal{I}_a$ *is a subfield of the following product field*

$$\mathcal{I}_a \subset \mathcal{F} \otimes \bigotimes_{b\in\mathbb{A}} \mathcal{U}_b \;, \quad \forall a \in \mathbb{A} \tag{1}$$

  and is called the information field *of the agent* a*.*

The *configuration space* is the product space (also called *hybrid space*, hence the $\mathbb{H}$ notation)

$$\mathbb{H} = \Omega \times \prod_{a\in\mathbb{A}} \mathbb{U}_a \;. \tag{2a}$$

The following product *configuration field* is a field over $\mathbb{H}$

$$\mathcal{H} = \mathcal{F} \otimes \bigotimes_{a\in\mathbb{A}} \mathcal{U}_a \;. \tag{2b}$$

A *policy*, of agent $a \in \mathbb{A}$ is a mapping

$$\lambda_a : (\mathbb{H}, \mathcal{H}) \to (\mathbb{U}_a, \mathcal{U}_a) \tag{3a}$$

from configurations to decisions, which satisfies the following measurability property

$$\lambda_a^{-1}(\mathcal{U}_a) \subset \mathcal{I}_a \;. \tag{3b}$$

Condition (3b) expresses the property that any policy of agent a may only depend upon the information $\mathcal{I}_a$ available to the agent. We denote by Λ_a the set of all policies of agent $a \in \mathbb{A}$. A *policy profile* λ is a collection of policies, one per agent $a \in \mathbb{A}$:

$$\lambda = (\lambda_a)_{a\in\mathbb{A}} \in \prod_{a\in\mathbb{A}} \Lambda_a \;. \tag{3c}$$

In what follows, we will need some notations. For any subset $B \subset \mathbb{A}$ of agents, we define

$$\mathcal{H}_B = \mathcal{F} \otimes \bigotimes_{b\in B} \mathcal{U}_b \otimes \bigotimes_{a\notin B} \{\emptyset, \mathbb{U}_a\} \subset \mathcal{H} \;, \tag{4a}$$

$$\lambda_B = (\lambda_b)_{b\in B} \in \Omega \times \prod_{b\in B} \mathbb{U}_b \;, \quad \forall \lambda = (\lambda_a)_{a\in\mathbb{A}} \in \prod_{a\in\mathbb{A}} \Lambda_a \;. \tag{4b}$$

2.2 Solvability and solution map

With any given policy profile $\lambda = (\lambda_a)_{a\in\mathbb{A}} \in \prod_{a\in\mathbb{A}} \Lambda_a$ we associate the set-valued mapping

$$\mathcal{M}_\lambda : \Omega \rightrightarrows \prod_{b\in\mathbb{A}} \mathbb{U}_b \;, \quad \omega \mapsto \Big\{ (u_b)_{b\in\mathbb{A}} \in \prod_{b\in\mathbb{A}} \mathbb{U}_b \;\Big|\; u_a = \lambda_a\big(\omega, (u_b)_{b\in\mathbb{A}}\big) \;, \quad \forall a \in \mathbb{A} \Big\} \;. \tag{5}$$

With this definition, we slightly reformulate below how Witsenhausen introduced solvability.

Definition 2. *([20, 21]) The* solvability property *holds true for the W-model of Definition 1 (or the W-model is* solvable*) when, for any policy profile* $\lambda = (\lambda_a)_{a\in\mathbb{A}} \in \prod_{a\in\mathbb{A}} \Lambda_a$*, the set-valued mapping* $\mathcal{M}_\lambda$ *in* (5) *is a mapping whose domain is* Ω*, that is, the cardinal of* $\mathcal{M}_\lambda(\omega)$ *is equal to one, for any state of nature* $\omega \in \Omega$.

Thus, under solvability property, for any state of nature $\omega \in \Omega$*, there exists one, and only one, decision profile* $(u_b)_{b\in\mathbb{A}} \in \prod_{b\in\mathbb{A}} \mathbb{U}_b$ *which is a solution of the* closed-loop equations

$$u_a = \lambda_a\big(\omega, (u_b)_{b\in\mathbb{A}}\big) \; , \quad \forall a \in \mathbb{A} \; . \tag{6a}$$

In this case, we define the solution map

$$S_\lambda : \Omega \to \mathbb{H} \; , \quad S_\lambda(\omega) = \big(\omega, M_\lambda(\omega)\big) \tag{6b}$$

where $M_\lambda(\omega)$ *is the unique element contained in the image set* $\mathcal{M}_\lambda(\omega)$ *that is, for all* $(u_b)_{b\in\mathbb{A}} \in \prod_{b\in\mathbb{A}} \mathbb{U}_b$*,* $M_\lambda(\omega) = (u_b)_{b\in\mathbb{A}} \iff \mathcal{M}_\lambda(\omega) = \{(u_b)_{b\in\mathbb{A}}\}$.

2.3 Causality

In his articles [20, 21], Witsenhausen introduces a notion of causality that relies on suitable configuration-orderings. Here, we introduce our own notations, as they make possible a compact formulation of the causality property.

Let $|\mathbb{A}|$ denote the cardinal of the set $\mathbb{A}$, that is, $|\mathbb{A}|$ is the number of agents. For $k \in \{1, \dots, |\mathbb{A}|\}$, let $\Sigma^k = \big\{\kappa : \{1, \dots, k\} \to \mathbb{A} \,\big|\, \kappa \text{ is an injection}\big\}$ denote the set of k-orderings, that is, injective mappings from $\{1, \dots, k\}$ to $\mathbb{A}$. The set $\Sigma^{|\mathbb{A}|}$ is the set of *total orderings* of agents in $\mathbb{A}$, that is, bijective mappings from $\{1, \dots, |\mathbb{A}|\}$ to $\mathbb{A}$. We define the *set of all partial orderings* by $\Sigma = \bigcup_{k\in\{0,\dots,|\mathbb{A}|\}} \Sigma^k$, where $\Sigma^0 = \{\emptyset\}$. For any $k \in \{1, \dots, |\mathbb{A}|\}$, any ordering $\kappa \in \Sigma^k$, and any integer $\ell \leq k$, $\kappa_{|\{1,\dots,\ell\}}$ is the restriction of the ordering κ to the first ℓ integers, and we introduce the mapping $\psi_k : \Sigma^{|\mathbb{A}|} \to \Sigma^k \; , \; \rho \mapsto \rho_{|\{1,\dots,k\}}$ which performs the restriction of any total ordering of $\mathbb{A}$ to $\{1, \dots, k\}$. For any $k \in \{1, \dots, |\mathbb{A}|\}$, and any partial k-ordering $\kappa \in \Sigma^k$, we define the *range* $\|\kappa\| = \big\{\kappa(1), \dots, \kappa(k)\big\} \subset \mathbb{A}$, the *cardinal* $|\kappa| = k \in \{1, \dots, |\mathbb{A}|\}$, the *last element* $\kappa^\star = \kappa(k) \in \mathbb{A}$, and the *restriction* $\kappa^- = \kappa_{|\{1,\dots,k-1\}} \in \Sigma^{k-1}$.

The next definition can be interpreted as follows. In a causal W-model, there exists a configuration-ordering with the following property: when an agent is called to play — as he is the last one in a partial ordering — what he knows cannot depend on decisions made by agents that are not his predecessors (in the range of the partial ordering under consideration).

Definition 3. *([20, 21]) A W-model (as in Definition 1) is* causal *if there exists (at least) one configuration-ordering* $\varphi : \mathbb{H} \to \Sigma^{|\mathbb{A}|}$ *with the property that*

$$\mathbb{H}^\varphi_\kappa \cap H \in \mathcal{H}_{\|\kappa^-\|} \; , \quad \forall H \in \mathcal{I}_{\kappa^\star} \; , \quad \forall \kappa \in \Sigma \; , \tag{7}$$

where the subset $\mathbb{H}^\varphi_\kappa \subset \mathbb{H}$ *of configurations is defined by (by convention, we put* $\mathbb{H}^\varphi_\emptyset = \mathbb{H}$*)*

$$\mathbb{H}^\varphi_\kappa = \big\{h \in \mathbb{H} \,\big|\, \psi_{|\kappa|}\big(\varphi(h)\big) = \kappa\big\} \; , \quad \forall \kappa \in \Sigma \; . \tag{8}$$

The set $\mathbb{H}^\varphi_\kappa$ contains all the configurations for which the agent $\kappa(1)$ is acting first, the agent $\kappa(2)$ is acting second,..., till the last agent $\kappa^\star = \kappa(|\kappa|)$ acting at stage $|\kappa|$. Hence, otherwise said, causality means that, once we know the first $|\kappa|$ agents, the information of the agent $\kappa^\star$ depends at most on the decisions of the agents in the range $\|\kappa^-\|$, as represented by the subfield (see Equation (4a))

$$\mathcal{H}_{\|\kappa^-\|} = \mathcal{F} \otimes \bigotimes_{a\in\|\kappa^-\|} \mathcal{U}_a \otimes \bigotimes_{b\notin\|\kappa^-\|} \{\emptyset, \mathbb{U}_b\} \subset \mathcal{H} \; . \tag{9}$$

In [20], Witsenhausen proves that causal W-models are solvable. He also shows that there exist solvable W-models that are not causal.

3 Generalizing causal graphical models concepts with the intrinsic model

We are now going to show how the basic notions of causal graphical models can be recovered and extended with the intrinsic model. Then, we will show an equivalence between three notions in the intrinsic model, namely blocking, directional separability and topological separability.

We recall that a *(binary) relation* $\mathfrak{R}$ on $\mathbb{A}$ is a subset $\mathfrak{R} \subset \mathbb{A}^2$ and that $a \,\mathfrak{R}\, b$ means $(a,b) \in \mathfrak{R}$. For any subset $B \subset \mathbb{A}$, the *(sub)diagonal relation* is $\Delta_B = \{(a,b) \in \mathbb{A}^2 \,|\, a = b \in B\}$ and the *diagonal relation* is $\Delta = \Delta_{\mathbb{A}}$. A *foreset* of a relation $\mathfrak{R}$ is any set of the form $\mathfrak{R}\, b = \{a \in \mathbb{A} \,|\, a \,\mathfrak{R}\, b\}$ or, by extension, of the form $\mathfrak{R}\, B = \{a \in \mathbb{A} \,|\, \exists b \in B \,,\ a \,\mathfrak{R}\, b\}$, where $B \subset \mathbb{A}$. The *opposite* or *complementary* $\mathfrak{R}^c$ of a binary relation $\mathfrak{R}$ is the relation $\mathfrak{R}^c = \mathbb{A}^2 \setminus \mathfrak{R}$, that is, defined by $a \,\mathfrak{R}^c\, b \iff \neg(a \,\mathfrak{R}\, b)$. The *converse* $\mathfrak{R}^{-1}$ of a binary relation $\mathfrak{R}$ is defined by $a \,\mathfrak{R}^{-1}\, b \iff b \,\mathfrak{R}\, a$ (and $\mathfrak{R}$ is *symmetric* if $\mathfrak{R}^{-1} = \mathfrak{R}$). The *composition* $\mathfrak{R}\mathfrak{R}'$ of two binary relations $\mathfrak{R}, \mathfrak{R}'$ is defined by $a(\mathfrak{R}\mathfrak{R}')b \iff \exists \delta \in \mathbb{A}$, $a \,\mathfrak{R}\, \delta$ and $\delta \,\mathfrak{R}'\, b$. The *transitive closure* of a binary relation $\mathfrak{R}$ is $\mathfrak{R}^\infty = \cup_{k=1}^{\infty} \mathfrak{R}^k$ (and $\mathfrak{R}$ is *transitive* if $\mathfrak{R}^\infty = \mathfrak{R}$) and the *reflexive and transitive closure* is $\mathfrak{R}^* = \mathfrak{R}^\infty \cup \Delta$.

3.1 Definition of conditional parentality in the intrinsic model

We suppose to be given a W-model as in Definition 1. Witsenhausen defines the *precedence relation* $\mathcal{P}$ on the set $\mathbb{A}$ of agents by

$$\mathcal{P}a = \bigcap_{B \in \mathbb{A}; \mathfrak{I}_a \subset \mathcal{H}_B} B \,, \ \forall a \in \mathbb{A} \quad \text{and} \quad b\mathcal{P}a \iff b \in \mathcal{P}a \,. \tag{10}$$

When the precedence relation $\mathcal{P}$ is acyclic, we recover a DAG. However, a W-model is much richer than the DAG it can induce. Instead of the limited precedence relation $\mathcal{P}$ in (10), we introduce a new and more flexible definition of parentality.

Definition 4. *For any subset $H \subset \mathbb{H}$ of configurations, and any subset $W \subset \mathbb{A}$ of agents, we set*

$$\mathcal{P}_{W,H}a = \bigcap_{B \in \mathbb{A}; \mathfrak{I}_a \cap H \subset \mathcal{H}_{B \cup W}} B \,, \ \forall a \in \mathbb{A} \,, \tag{11a}$$

and we define the (conditional) parental relation $\mathcal{P}_{W,H}$ *on $\mathbb{A}$ (w.r.t. (W,H)) by*

$$b \,\mathcal{P}_{W,H}\, a \iff b \in \mathcal{P}_{W,H}a \,, \ \forall (a,b) \in \mathbb{A}^2 \,. \tag{11b}$$

Thus, when $b\mathcal{P}_{W,H}a$, it means that the information available to agent a, on the subset $H \subset \mathbb{H}$ of configurations, necessarily involves the decisions of the agent b and, possibly of the agents in W. Witsenhausen's precedence relation $\mathcal{P}$ is the special case $\mathcal{P}_{\emptyset,H}$. We have that $\mathcal{P}_{W,H}\mathbb{A} \subset W^c$ by construction, and it can be shown that $\mathcal{P}_{W,H} = \Delta_{W^c}\mathcal{P}_{\emptyset,H}$.

3.2 Definitions of conditional blocking, and of directional and topological separation in the intrinsic model

We take inspiration from Pearl to define the two following notions of conditional blocking and of conditional directional separation. Then, we introduce a suitable topology on the set of agents, and we define a new notion of conditional topological separation.

3.2.1 Definition of conditional blocking

We define *conditional blocking* using the language of relations.

Definition 5. *We suppose given a subset $H \subset \mathbb{H}$ of configurations, and a subset $W \subset \mathbb{A}$ of agents.*

We define the (conditional) genealogical relation $\mathcal{A}_{W,H}$ *(w.r.t. (W,H)) as the transitive and reflexive closure of the conditional parental relation $\mathcal{P}_{W,H}$ in Definition 4, that is,*

$$\mathcal{A}_{W,H} = \mathcal{P}_{W,H}^{\infty} \cup \Delta = \mathcal{P}_{W,H}^{*} , \tag{12a}$$

the (conditional) common cause relation $\mathcal{K}_{W,H}$ *(w.r.t. (W,H)) as the symmetric relation*

$$\mathcal{K}_{W,H} = \mathcal{A}_{W,H}^{-1} \Delta_{W^c} \mathcal{A}_{W,H} , \tag{12b}$$

the (conditional) connection relation $\mathcal{C}_{W,H}$ *(w.r.t. (W,H)) as the symmetric relation*

$$\mathcal{C}_{W,H} = \Big(\mathcal{K}_{W,H} \big(\Delta_W \mathcal{K}_{W,H} \Delta_W \cup \Delta_W \big)^{\infty} \mathcal{K}_{W,H} \Big) \cup \mathcal{K}_{W,H} , \tag{12c}$$

and finally the (conditional) blocking relation $\mathcal{B}_{W,H}$ *(w.r.t. (W,H)) as the complementary relation*

$$\mathcal{B}_{W,H} = \mathcal{C}_{W,H}^{c} . \tag{12d}$$

3.2.2 Definition of conditional directional separation

Now we propose a reinterpretation of Dawid's notation [4] to define *conditional directional separation* in our context.

Definition 6. *We suppose given a subset $H \subset \mathbb{H}$ of configurations, and a subset $W \subset \mathbb{A}$ of agents.*

1. *Using the precedence relation $\mathcal{P}_{\emptyset,H}$, which is a subset of $\mathbb{A}\times\mathbb{A}$, we introduce the oriented graph $\mathcal{G}_H = (\mathbb{A}, \mathcal{P}_{\emptyset,H})$ and the undirected graph $\hat{\mathcal{G}}_H = (\mathbb{A}, \mathcal{P}_{\emptyset,H} \cup \mathcal{P}_{\emptyset,H}^{-1})$ (obtained by considering the symmetric closure of the relation $\mathcal{P}_{\emptyset,H}$). The nodes of the graphs $\mathcal{G}_H$ and $\hat{\mathcal{G}}_H$ are the agents in $\mathbb{A}$. The arcs of the graph $\mathcal{G}_H$ are the couples of nodes in $\mathcal{P}_{\emptyset,H} \subset \mathbb{A}\times\mathbb{A}$. The arcs of the graph $\hat{\mathcal{G}}_H$ are the couples of nodes in $\mathcal{P}_{\emptyset,H} \cup \mathcal{P}_{\emptyset,H}^{-1} \subset \mathbb{A}\times\mathbb{A}$.*

2. *Let b, $c \in \mathbb{A}$ be two agents. Let π be a path in the undirected graph $\hat{\mathcal{G}}_H$ that joins b and c (considered as two nodes in the set $\mathbb{A}$ of nodes of the undirected graph $\hat{\mathcal{G}}_H$). We say that the path π* directionally separates (w.r.t. (W,H)) *the two nodes b and c if all the paths in the oriented graph $\mathcal{G}_H$ that give rise to the path π in the undirected graph $\hat{\mathcal{G}}_H$ contain a subpath of three consecutive nodes b', w, $c' \in \mathbb{A}$ satisfying one of the following four conditions:*

   - $w \in W$ *and* $b'\mathcal{P}_{\emptyset,H}w$ *and* $w\mathcal{P}_{\emptyset,H}c'$,
   - $w \in W$ *and* $b'\mathcal{P}_{\emptyset,H}^{-1}w$ *and* $w\mathcal{P}_{\emptyset,H}^{-1}c'$,
   - $w \in W$ *and* $b'\mathcal{P}_{\emptyset,H}^{-1}w$ *and* $w\mathcal{P}_{\emptyset,H}c'$,
   - $w \notin \mathcal{P}_{\emptyset,H}^{*} W$ *and* $b'\mathcal{P}_{\emptyset,H}w$ *and* $w\mathcal{P}_{\emptyset,H}^{-1}c'$.

3. *Let b, c be two agents in $\mathbb{A}$. We say that the two agents b and c are (conditionally)* directionally separated *(w.r.t. (W,H)), denoted by $b \perp\!\!\!\perp_d c \mid (W,H)$, if any path joining the two nodes b and c in the undirected graph $\hat{\mathcal{G}}_H$ directionally separates the two nodes.*

4. *For two subsets $B, C \subset \mathbb{A}$ such that*

   $$B \cap C = \emptyset , \quad B \cap W = \emptyset , \quad C \cap W = \emptyset , \tag{13a}$$

   we say that B and C are (conditionally) directionally separated *(w.r.t. (W,H)), denoted by $B \perp\!\!\!\perp_d C \mid (W,H)$, when for any agent $b \in B$ and any agent $c \in C$, the two agents b and c are (conditionally)* directionally separated *(w.r.t. (W,H)).*

3.2.3 Definition of conditional topological separation

We introduce a suitable topology on the set of agents, extending the topology introduced in [21], and we define a new notion of *conditional topological separation*.

Proposition 7. *The following subset $\mathcal{T}_{W,H}$ of $2^{\mathbb{A}}$ is a topology on the set $\mathbb{A}$ of agents:*

$$\mathcal{T}_{W,H} = \{B \subset \mathbb{A} \mid \mathcal{A}_{W,H}(\mathbb{A} \setminus B) \subset \mathbb{A} \setminus B\} , \tag{14}$$

where the genealogical relation $\mathcal{A}_{W,H}$ has been defined in (12a)*. In this topology, the subset W is open, a subset $C \subset \mathbb{A}$ is closed iff $\mathcal{A}_{W,H}C \subset C$ iff $\mathcal{A}_{W,H}C = C$, and the topological closure $\overline{B}^{W,H}$ of a subset $B \subset \mathbb{A}$ is the foreset*

$$\overline{B}^{W,H} = \mathcal{A}_{W,H}B . \tag{15}$$

For any subsets $B \subset \mathbb{A}$ and $B_j \subset \mathbb{A}$, $j = 1, \ldots, n$, we write $B_1 \sqcup \cdots \sqcup B_n = B$ when we have both $B_j \cap B_k = \emptyset$ for all $j \neq k$ and $B_1 \cup \cdots \cup B_n = B$.

Definition 8. *We suppose given a subset $H \subset \mathbb{H}$ of configurations, and a subset $W \subset \mathbb{A}$ of agents. We say that two subsets $B, C \subset \mathbb{A}$ are (conditionally)* topologically separated *(w.r.t. (W, H)), denoted by $B \underset{t}{\perp\!\!\!\perp} C \mid (W, H)$, if there exists $W_B, W_C \subset W$ such that*

$$W_B \sqcup W_C = W \text{ and } \overline{B \cup W_B}^{W,H} \cap \overline{C \cup W_C}^{W,H} = \emptyset . \tag{16}$$

When $B = \{b\}$ and $C = \{c\}$, we say that b and c are topologically separated, denoted by $b \underset{t}{\perp\!\!\!\perp} c \mid (W, H)$, as a shorthand for $\{b\}$ and $\{c\}$ are topologically separated.

When $B, C \subset \mathbb{A}$ are topologically separated, we necessarily have that $B \cap C = \emptyset$.

3.3 Equivalence between blocking, directional and topological separation in the intrinsic model

With the above definitions, we now show that the notions of conditional blocking, conditional directional separability and conditional topological separation are equivalent.

Theorem 9. *For any pair of agents $b, c \in \mathbb{A}$ such that*

$$b \neq c , \quad b \notin W , \quad c \notin W , \tag{17}$$

the following statements are equivalent: the two agents b and c are

1. *conditionally blocked (w.r.t. (W, H)), that is, $b \, \mathcal{B}_{W,H} \, c$,*
2. *conditionally directionally separated (w.r.t. (W, H)), that is, $b \underset{d}{\perp\!\!\!\perp} c \mid (W, H)$,*
3. *conditionally topologically separated (w.r.t. (W, H)), that is, $b \underset{t}{\perp\!\!\!\perp} c \mid (W, H)$.*

Hence, in any of the three equivalent cases, we will say that the two agents b and c are *conditionally separated* (w.r.t. (W, H)), denoted by $b \perp\!\!\!\perp c \mid (W, H)$, and the same for two subsets $B, C \subset \mathbb{A}$ satisfying (13a).

4 Conditional separation, factorization and do-calculus in intrinsic models

In this section, where we deal with probability, we consider a finite W-model (as in Definition 1), that is, where all sets are finite, to avoid technical measurability issues. Moreover, we suppose that the set Ω of states of Nature, and its field $\mathcal{F}$ have the following product form:

$$\Omega = \prod_{a \in \mathbb{A}} \Omega_a , \quad \mathcal{F} = \bigotimes_{a \in \mathbb{A}} \mathcal{F}_a . \tag{18a}$$

For any subset $B \subset \mathbb{A}$ of agents, we denote

$$\Omega_B = \prod_{b \in B} \Omega_b \; , \;\; \mathcal{F}_B = \bigotimes_{b \in B} \mathcal{F}_b \; , \;\; \mathbb{U}_B = \prod_{b \in B} \mathbb{U}_b \; , \;\; \mathcal{U}_B = \bigotimes_{b \in B} \mathcal{U}_b \; , \tag{18b}$$

and we denote by π_B the projection from $\mathbb{H}$ to its factors in B as follows:

$$\pi_B : \mathbb{H} = \Omega_B \times \Omega_{B^c} \times \mathbb{U}_B \times \mathbb{U}_{B^c} \to \mathbb{U}_B \; . \tag{18c}$$

4.1 Conditional separation implies factorization

We are now going to show how conditional separation induces a factorization of the solution map.

Theorem 10. *We consider a solvable W-model, where the set Ω of states of Nature has the product form* (18a)*, where each information field $\mathcal{I}_a$ in* (1) *is such that*

$$\mathcal{I}_a \subset \mathcal{F}_a \otimes \bigotimes_{b \neq a} \{\emptyset, \Omega_b\} \otimes \bigotimes_{c \in \mathbb{A}} \mathcal{U}_c \; , \quad \forall a \in \mathbb{A} \; . \tag{19}$$

We also consider a policy profile $\lambda = (\lambda_a)_{a \in \mathbb{A}} \in \prod_{a \in \mathbb{A}} \Lambda_a$, a subset $H \subset \mathbb{H}$ of configurations, and Y, W and Z three disjoint subsets of $\mathbb{A}$, such that (see the comment following Theorem 9)

$$Y \perp\!\!\!\perp Z \mid (W, H) \; . \tag{20}$$

Then, there exist five subsets Y', Z', W_Y, W_Z, $\langle\emptyset\rangle \subset \mathbb{A}$ such that

$$\mathbb{A} = \widetilde{Y} \sqcup \tilde{Z} \sqcup \widetilde{\emptyset} \quad \textit{where} \quad \widetilde{Y} = Y \sqcup Y' \sqcup W_Y \; , \;\; \tilde{Z} = Z \sqcup Z' \sqcup W_Z \; , \;\; W = W_Y \sqcup W_Z \; , \tag{21a}$$

and there exist three mappings (reduced solution maps)

$$\widetilde{M}_{\lambda_{\widetilde{Y}}} : \Omega_{\widetilde{Y}} \times \mathbb{U}_{W_Z} \to \mathbb{U}_{\widetilde{Y}} \; , \;\; \widetilde{M}_{\lambda_{\widetilde{Z}}} : \Omega_{\widetilde{Z}} \times \mathbb{U}_{W_Y} \to \mathbb{U}_{\widetilde{Z}} \; , \;\; \widetilde{M}_{\lambda_{\langle\emptyset\rangle}} : \Omega_{\langle\emptyset\rangle} \times \mathbb{U}_{\widetilde{Y} \cup \widetilde{Z}} \to \mathbb{U}_{\langle\emptyset\rangle} \tag{21b}$$

such that the solution map (6b) *splits in three factors as follows: $\forall \omega \in S_\lambda^{-1}(H)$, we have that*

$$M_\lambda(\omega) = \widetilde{M}_{\lambda_{\widetilde{Y}}}\Big(\omega_{\widetilde{Y}}, \lambda_{W_Z}\big(S_\lambda(\omega)\big)\Big) \times \widetilde{M}_{\lambda_{\widetilde{Z}}}\Big(\omega_{\widetilde{Z}}, \lambda_{W_Y}\big(S_\lambda(\omega)\big)\Big) \times \widetilde{M}_{\lambda_{\langle\emptyset\rangle}}\Big(\omega_{\langle\emptyset\rangle}, \lambda_{\widetilde{Y} \cup \widetilde{Z}}\big(S_\lambda(\omega)\big)\Big) \; . \tag{21c}$$

More precisely, with the notation (18c), Equation (21c) has to be understood as $\pi_{\widetilde{Y}}\big(M_\lambda(\omega)\big) = \widetilde{M}_{\lambda_{\widetilde{Y}}}\Big(\omega_{\widetilde{Y}}, \lambda_{W_Z}\big(S_\lambda(\omega)\big)\Big)$, $\pi_{\widetilde{Z}}\big(M_\lambda(\omega)\big) = \widetilde{M}_{\lambda_{\widetilde{Z}}}\Big(\omega_{\widetilde{Z}}, \lambda_{W_Y}\big(S_\lambda(\omega)\big)\Big)$, and $\pi_{\langle\emptyset\rangle}\big(M_\lambda(\omega)\big) = \widetilde{M}_{\lambda_{\langle\emptyset\rangle}}\Big(\omega_{\langle\emptyset\rangle}, \lambda_{\widetilde{Y} \cup \widetilde{Z}}\big(S_\lambda(\omega)\big)\Big)$.

4.2 Probabilistic implications and Do-calculus in intrinsic models

Theorem 11. *We suppose that the assumptions of Theorem 10 are satisfied, and that all fields are σ-fields. Moreover, we suppose that the set Ω in* (18a) *is equipped with a probability $\mathbb{P} = \bigotimes_{a \in \mathbb{A}} \mathbb{P}_a$ where each $\mathbb{P}_a$ is a probability on $(\Omega_a, \mathcal{F}_a)$.*

We define the following pushforward probability $\mathbb{Q}_\lambda$ on $(\mathbb{H}, \mathcal{H})$ by

$$\mathbb{Q}_\lambda = \mathbb{P} \circ S_\lambda^{-1} \; . \tag{22}$$

Then, $\big(\mathbb{H}, \mathcal{H}, \mathbb{Q}_\lambda\big)$ is a probability space, and the two projections $\pi_{\overline{Y}^{W,H}} : (\mathbb{H}, \mathcal{H}) \to (\mathbb{U}_{\overline{Y}^{W,H}}, \mathcal{U}_{\overline{Y}^{W,H}})$ and $\pi_{\overline{Z}^{W,H}} : (\mathbb{H}, \mathcal{H}) \to (\mathbb{U}_{\overline{Z}^{W,H}}, \mathcal{U}_{\overline{Z}^{W,H}})$ as in (18c) *are independent under $\mathbb{Q}_\lambda$, conditionally on the subset $H \subset \mathbb{H}$ and on the projection $\pi_W : (\mathbb{H}, \mathcal{H}) \to (\mathbb{U}_W, \mathcal{U}_W)$.*

4.3 Intervention variables, Do-calculus

We now introduce the possibility to intervene on a variable. We can encode this possibility in the model using a simple procedure. Suppose we are interested in an intervention policy profile $\hat{\lambda}_Z$ for a subset $Z \subset \mathbb{A}$ of agents. For this purpose, we consider a new family of fields $\hat{\mathcal{I}}_z \subset \mathcal{H}$, for $z \in Z$, as in (1), and we suppose that $\hat{\lambda}_Z$ is $\hat{\mathcal{I}}_z$-measurable, for any $z \in Z$, as in (3b). Then, we enrich the W-model as follows (detailed in the companion paper): (1) we introduce a new *intervention agent* I, equipped with $\Omega_I = \{0,1\}$ and $\mathbb{U}_I = \{0,1\}$, and who only has access to her/his private information in Ω_I; (2) we straightforwardly adapt the information fields for $\mathbb{A} \setminus (Z \cup I)$ and the probability $\mathbb{P}$; (3) we replace the information field $\mathcal{I}_z$ by $\{0\} \otimes \mathcal{I}_z \cup \{1\} \otimes \hat{\mathcal{I}}_z$, for $z \in Z$.

Theorem 12. *Under the assumptions of Theorem 11, we have that the projection π_Y has the same conditional distribution under $\mathbb{Q}_\lambda$, whether the conditioning is w.r.t. the subset $H \subset \mathbb{H}$, the projection π_W and the projection $\pi_{\overline{Z}^{W,H}}$, or is only w.r.t. the subset $H \subset \mathbb{H}$ and the projection π_W.*

This rule subsumes Pearl's do-calculus.

We have proved, loosely speaking, that

$$Y \perp\!\!\!\perp Z \mid (W,H) \implies \mathbb{Q}_\lambda(h_Y | h_W, h_{\overline{Z}^{W,H}}, H) = \mathbb{Q}_\lambda(h_Y | h_W, H) \ . \tag{23}$$

We stress the conciseness of Theorem 12 — permitted by the notions introduced in this paper — that implies the three rules of Pearl and also the following two recent results.

Proposition 13. *Rule 1 from [17] and Theorem 1 from [2] can be deduced from Theorem 12.*

5 Discussion

In this paper, we simplify and generalize the do-calculus by leveraging the concepts of information field and solution map. The do-calculus is reduced to one rule. Causality is presented as a property of the W-model, and is not encoded by design. The intrinsic model is richer than DAGs, and allows for cycles, conditioning, and even noncausality as we only suppose the more general assumption of solvability, while DAGs warrant the modeling of situations that do not possess a fixed causal ordering [7]. Statistical independence follows from the factorization property of the solution map implied by separation (related to a form of informational independence). We underline that the results come from the information structure, not the probability. Also, because our approach is not based on graphical models, our work provides a new proof of Pearl's original result.

Also, it is notable that one of Witsenhausen's motivation was Kuhn's extensive games [8, 18], where information plays a central role, and the question of information modeling is still debated [1, 6].

Further work includes drawing connections with other research programs, such as Proposition 13 or questions related to identification [14, 15, 16], using the framework developed in this paper.

Acknowledgments

We would like to thank David Rohde for his comments and suggestions.

References

[1] C. Alós-Ferrer and K. Ritzberger. *The theory of extensive form games*. Springer Series in Game Theory. Springer-Verlag, Berlin, 2016.

[2] J. Correa and E. Bareinboim. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *AAAI 2020 : The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.

[3] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media, 2006.

[4] A. P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):1–15, 1979.

[5] J. Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

[6] B. Heymann, M. De Lara, and J.-P. Chancelier. Kuhn's equivalence theorem for games in intrinsic form, 2020. preprint.

[7] S. Itani, M. Ohannessian, K. Sachs, G. P. Nolan, and M. A. Dahleh. Structure learning in causal cyclic networks. In *Causality: Objectives and assessment*, pages 165–176, 2010.

[8] H. W. Kuhn. Extensive games and the problem of information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, pages 193–216. Princeton University Press, Princeton, 2016.

[9] F. Lattimore and D. Rohde. Replacing the do-calculus with Bayes rule. *arXiv preprint arXiv:1906.07125*, 2019.

[10] J. Pearl. The mathematics of causal relations. *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures (P. Shrout, K. Keyes and K. Ornstein, eds.). Oxford University Press, Corvallis, OR*, pages 47–65, 2010.

[11] J. Pearl. *Causality: Models, reasoning, and inference*. Cambridge University Press, second edition edition, 2011.

[12] J. Pearl and D. Mackenzie. *The book of Why: the new science of cause and effect*. Basic Books, 2018.

[13] J. Peters, D. Janzing, and B. Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

[14] I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06*, volume 2, pages 1219–1226. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[15] I. Shpitser and J. Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9(Sep):1941–1979, 2008.

[16] S. Tikka, A. Hyttinen, and J. Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *arXiv preprint arXiv:1902.01073*, 2019.

[17] S. Tikka, A. Hyttinen, and J. Karvanen. Identifying causal effects via context-specific independence relations. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 2804–2814. Curran Associates, Inc., 2019.

[18] J. von Neumann and O. Morgenstern. *Theory of games and economic behaviour*. Princeton University Press, 1944.

[19] J. Winn. Causality with gates. In *Artificial Intelligence and Statistics*, pages 1314–1322, 2012.

[20] H. S. Witsenhausen. On information structures, feedback and causality. *SIAM J. Control*, 9(2):149–160, May 1971.

[21] H. S. Witsenhausen. The intrinsic model for discrete stochastic control: Some open problems. In A. Bensoussan and J. L. Lions, editors, *Control Theory, Numerical Methods and Computer Systems Modelling*, volume 107 of *Lecture Notes in Economics and Mathematical Systems*, pages 322–335. Springer-Verlag, 1975.